DOCUMENT RESUME

ED 422 402 TM 028 964

AUTHOR Lin, Chow-Hong; Davenport, Ernest C., Jr.
TITLE Comparison on a Robust Reweighted Least Squares Procedure and Four Other Regression Techniques.
PUB DATE 1998-04-00
NOTE 33p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Comparative Analysis; *Least Squares Statistics; *Regression (Statistics); *Robustness (Statistics); Tables (Data)
IDENTIFIERS *Weighting (Statistical)

ABSTRACT

The iterative expanding reweighted least squares (IERLS), a weighted-least-squares robust regression procedure, is developed and compared with Huber's M estimator (P. Huber, 1973), the Least Absolute Deviations, the Least Median of Squares, and the Ordinary Least Squares. Data with various outliers contamination were simulated and analyzed. The results suggest several things. First, any procedure is not superior under all circumstances. A procedure that reduces the bias on regression weight estimates might at the same time increase their stability. The Least Median of Squares (LMS) is found to be much more efficient than the other three robust techniques with two regressors, but less efficient when there is only one predictor. Second, the magnitude of differences between estimates by each procedure and the true parameters are smallest for LMS (n=50) or the proposed procedure (n=100). Third, the proposed reweighted approach appears to excel all but the LMS procedure with complex data with higher dimensions and contamination. If both computing time and protection against outliers are important, the proposed reweighted technique seems to be a more plausible robust procedure than the three other robust procedures studied in this paper. Appendixes illustrate the relationship between the mean squared error and variances, conditions for simple and multiple regression, parameter estimates studied in the simulation, and the annotated Statistical Package for the Social Sciences program for the IERLS procedure. (Contains 7 tables and 10 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 422 402

Comparison on a Robust Reweighted Least Squares Procedure and Four Other Regression Techniques

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY HAS BEEN GRANTED BY

Chowhong Lin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Chow-Hong Lin and Ernest C. Davenport, Jr.

Department of Educational Psychology

University of Minnesota

Requests for reprints should be addressed to Chow-Hong Lin, Department of Educational Psychology, University of Minnesota, 178 Pillsbury Drive, S.E., Minneapolis, MN 55455.

Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, California, April 13-17, 1998.

TM028964

Abstract

A weighted-least-squares based robust regression procedure is developed and compared with Huber's M estimator, the Least Absolute Deviations, the Least Median of squares, and the Ordinary least squares. Data with various outliers contamination were simulated and analyzed. The results suggest several things. First, any procedure is not superior under all circumstances. A procedure which reduces the bias on regression weight estimates might at the same time increase their instability. LMS is found to be much efficient than the other three robust techniques with two regressors, but less efficient when there is only one predictor. Second, the magnitude of differences between estimates by each procedure and the true parameters are smallest for LMS (n=50) or the proposed procedure (n=100). Third, the proposed reweighted approach appears to excel all but the LMS procedure with complex data with higher dimensions and contamination. If both computing time and protection against outliers are important, the proposed reweighted technique seems to be a more plausible robust procedure than the three other robust procedures studied in this paper.

Introduction

A robust regression procedure can serve one of two purposes. First, to identify outliers. Second, to estimate the true regression coefficients. Oftentimes, a robust regression procedure can accomplish both with a weighted least squares approach. Data points receive differential weights based on some criteria, and those that have small weights could be identified as potential outliers.

This paper is an extension of Lin and Davenport (1997), who proposed an iterative procedure in an attempt to cope with problems caused by the presence of outliers in the data. They used five existing data sets, and found the iterative procedure useful in identifying outliers. This study applied their iterative method and three other robust regression procedures: the Least Median of Squares (Roûsseeuw, 1984; Rousseeuw and Leroy, 1987), the Least Absolute Deviations (Bloomfield and Steiger, 1983), and Huber's M-estimator (with the tuning constant set to 1.5, Huber, 1973), to simulated data; and compared each procedure's capacity to recover underlying models when outliers exist in the data. The Ordinary Least Squares procedure was used as a basis for comparison.

Both bivariate and "tri"-variate data were generated based on the models to be described in a later section. Outliers were then simulated based on three criterion: the percentage of outliers, the magnitude of outlying, and the variable(s) in which outliers were outlying. This resulted in 55 and 127 distinct conditions for simple regression and multiple regression with p=2, respectively.

Four hundred replications were performed within each condition, and sampling distributions of each of the beta estimates in each procedure are therefore available. Four criterion were used to compare the performance of beta estimates of each procedure. First is the efficiency relative to OLS, Ordinary Least Squares. The second is the MSE (Mean Squared Errors) ratios relative to OLS. The third is the bias, defined as the absolute difference between the average beta estimates over 400 replications and the true

parameter values. Finally, another measure of bias, defined as the ratio of MSE and variance of beta estimates, was contrasted. In addition, this study also provides a comparison of the execution time required for each procedure.

The Iterative Expanding Reweighted Least Squares (IERLS) Procedure

An iteratively reweighted least squares (IRLS) procedure accomplishes the estimation process in three steps, iterating between steps 2 and 3 until a certain influence function (overall error) is minimized. First, obtain initial estimates of the regression coefficients and standard error based on some "robust" procedure. Second, perform a weighted least squares procedure based on weights that are some inverse function of the (un)scaled residuals from the previous step. Calculate the residual using the WLS solution. Third, compare the overall error ("the influence function") with certain criteria to determine whether the relative accuracy (between two consecutive runs) has been achieved. If not achieved, then update the weights using the residuals from step 2 and loop through steps 2 and 3, again until a relative accuracy is obtained.

The Iterative Expanding Reweighted Least Squares procedure utilizes the IRLS. It calculates first the distance between each data point to the sample centroid, and selects the half-sample data points that are closest to the centroid. It then performs an OLS regression on this subset of data points and obtains an initial regression coefficient estimates.

The weights are determined in two stages. After the initial regression coefficients are obtained, the residual of every data point in the full sample is calculated, and points are sorted a second time, this time based on the magnitude of residuals to the fitted line/surface. Each of the first 50% data points with the lowest residuals has a weight of 1, and the rest has the weight

$$1/(1+|r_i|) \qquad (1)$$

where $|r_i|$ is the absolute value of the i-th residual, given that the i-th data point is not one of the 50 % data points with the smallest residuals

This is the first-stage weight. After the residual of each data point (for all points) is calculated, the basis will now expand to have 50% of the full sample plus 1 data points with the lowest residuals. (Note that this second sort is performed to avoid possible mis-classification of data points that are far away from the centroid but are not outliers.) A weighted least squares is then performed to the included points, with the weight from the first stage rescaled such that the sum of rescaled weights is the number of data points included in each iteration. If w_{old} is the weight from stage 1, then

$$w_{new} = \frac{n_k \times w_{old}}{\sum w_{old}} \qquad (2)$$

where n_k is the number of included data points at iteration k

is the stage two weight.

The weighted least squares is performed with the stage two weight, and the regression coefficients are obtained. The program then proceeds to take in the data point with the next smallest residual, and the stage one weights of data points included in the current expanded set, are rescaled (since the number of included data points is changed) to obtain the updated stage two weights. An WLS is then executed on the expanded subset of data, and regression coefficients again obtained.

This iteration procedure is repeated until all data points are in the subset. Then the subset is just the full sample. The overall regression coefficients are the weighted averages of the regression coefficients at each iteration, weighted by n_k for all k

Algorithm for the proposed Iterative Expanding Reweighted Least Squares

Step 1 Input the data matrix, which is the concatenation of both the dependent and independent variables. In matrix form, write

$$Z=[X|Y]$$

Here the dimensionality of the Z matrix is n rows by p (without an intercept) or $p + 1$ (with an intercept) columns, where p is the number of predictors.

Step 2 Calculate the distance measure for observation i (D_i) by using

$$D_i = \left(Z_{ij} - \bar{Z}_j\right)' \left(s^2\right)^{-1} \left(Z_{ij} - \bar{Z}\right), \text{ for } i = 1 \text{ to } n \qquad (3)$$

Step 3 Sort observations in ascending order on the absolute value of D_i.

Step 4 Select the first half of the points with the shortest distance measures as the basis and perform an Ordinary Least Squares regression.

Step 5 Calculate the unstandardized residuals for all data points based on the regression coefficients obtained in step 4.

Step 6 Sort all observations in ascending order of magnitudes of residuals from step 5.

Step 7 Calculate weight associated with each data point via equation (1). Note that the first 50% of the data points will have an initial weight of 1.

Step 8 Select the first $n/2 + 1$ or $[(n + 1)/2 + 1]$ points, depending on number of observations. This is the initial working data set.

Step 9 Rescale weights obtained in Step 7, so that the sum of rescaled weights is the number of observations used at that iteration. Perform a weighted least squares regression based on the rescaled weights. Thus, when points are added into the initial working data, the rescaled weight for the initial half sample observations will be a constant larger than unity. Those points added later will have a weight

less than unity. (The sample size at the k-th step will be one larger than that at the previous step, and the sum of weights will equal the sample size at each iteration.) Store the regression coefficients.

Step 10. If all observations are in the working data set, go to Step 12. Otherwise, include the observation that is next on the list (with the next smallest error) to the working data set.

Step 11. Go to step 9.

Step 12. Calculate the weighted averages of each regression coefficient using the sample size at each iteration as the weight.

Step 13. Stop.

Appendix 5 is an annotated SPSS program to perform IERLS. Please note that in the program, the standard errors of the estimates were also calculated. However, it was found later that those standard errors were too liberal. The MS-Fortran version of the program is available from the first author.

Simulation of the Data

Sample size

- Two sample sizes were used in this study: 50 and 100.

Generation of the data:

Three variables were created following the conditions below:

$$X_1 \sim N(4,10)$$
$$X_2 \sim N(2,6)$$
$$X_3 \sim N(0,1)$$
$$Y = X_1 + X_2 + X_3 \qquad (4)$$

(For simple regression, X_2 was regressed on X_1, for multiple regression, both X_1 and X_2 were regressed on Y. X_3 was added to Y to simulate the random variation in Y, over and above X_1 and X_2.)

The creation of random normal numbers is as follows. Based on the central limit theorem, that even though the population distribution is skewed (in this study, a uniform distribution is used), given an adequate sample size n and a sampling procedure (random), the sampling distribution of the means will tend to approximate a normal distribution, with the expected mean being the population mean. One can thus standardize the sum of these numbers by subtracting out the expected mean from each number, and then dividing the difference by the squared root of the expected variance. The distribution of thus derived "standardized" values should likely conform to a standard normal distribution.

It is known that for a uniform distribution between two numbers a, b ($b > a$), the mean is $\frac{b+a}{2}$, and the variance is $\frac{b-a}{12}$ (Lewis and Orav, 1989; Tietjen, 1986). Thus, if one draws m numbers independently from a $U(0,1)$ distribution, the *expected* sum and variance of the m numbers will be

$$m \times \frac{1+0}{2} = \frac{m}{2} \quad \text{and} \tag{5a}$$

$$m \times \frac{1-0}{12} = \frac{m}{12}, \quad \text{respectively} \tag{5b}$$

The standard error is therefore $\sqrt{\frac{m}{12}}$, the squared root of $\frac{m}{12}$. If these m numbers are summed and standardized by the sum of the expected mean and standard error, the standardized value would thus represent a random observation from a standard normal distribution, if m is large enough. Since the sample size was either 50 or 100, such "sampling" process would repeat for either 50 or 100 times.

Woodward and Overall (1974) used $m = 12$ in their random number generation subroutine. However, their program requires user input of a vector of 100 z-scores, from -2.68 to 2.68, for converting uniform random numbers to normal random numbers.

The expected pairwise correlation between the three variables (X_1, X_2, and X_3) was zero, since each variable was generated independently. In passing, it can be noted that the *expected* correlations between Y and X_1, X_2, and X_3 are 0.14, 0.19, and 0.06, respectively. The use of uncorrelated predictors for a regression problem was justified by the fact that, through an appropriate transformation, uncorrelated variables can have any desired correlation, refer to Appendix 1. (Chatterjee and Hadi, 1988; Banchoff and Wermer, 1992)

Planting of outliers

Outliers can influence parameter estimates through at least three different ways that are investigated in this study: the relative percentage of outliers present (some data sets have more outliers than others), the dispersion of the outliers (some data sets have outliers that lie very far away from the bulk of the data), and the dimension of the outliers (some outliers are outliers in terms of the independent variable(s), others are in terms of the dependent variable, still others are in terms of both dependent and independent variables).

percentage of outliers, dispersion of outliers, and dimension of outliers:

Four different percentages of data points were then transformed to become "outliers": 0%, 5%, 15%, and 25%. The transformed data points were multiplied by either 3, -3, 6, -6, 9, -9, relative to the initial values for X_1, X_2 and/or Y. If more than one variable for any observation is chosen to be "outlying", all variable values for that observation are multiplied by the same value. One additional set of conditions was added where the outlying dimension is independently determined. That is, which variable of which observation should be outlying was decided separately, instead of simultaneously.

Therefore, 72 conditions (for simple regression) and 168 conditions (for multiple regression with two predictors) resulted (see Appendices 2 and 3 for descriptions). Four hundred replications were made, thus forming a sampling distribution with 400 estimated values for each of the regression parameter estimates. These are summarized in Appendix 4.

The population regression models were:

$$\text{Model 1:} \quad X_2 = 2 + 0 \times X_1 + \varepsilon \qquad \text{(simple regression)} \qquad (6a)$$

$$\text{Model 2:} \quad Y = 0 + 1 \times X_1 + 1 \times X_2 + \varepsilon \qquad \text{(multiple regression)} \qquad (6b)$$

Therefore, the regression coefficients for the population were:

For simple regression:

$\beta_0 = 2, \beta_1 = 0$

For multiple regression:

$\beta_0 = 0, \beta_1 = 1, \beta_2 = 1$

Since the parameters are known, detailed comparison can be made. The mean of each sampling distribution, as well as the MSE and bias as defined previously are compared among these procedures. Parameter estimates of a variation of the IERLS procedure is also compared. Specifically, IERLS uses the weighted average of regression coefficients from each iteration as the final estimates. The variation, however, uses the simple unweighted average from each iteration as the final estimates.

Major Results and Discussions

The procedures compared in this study are represented by the following:

HUB Huber's M-estimate (with $k = 1.5$)

LAD	Least Absolute Deviations
LMS	Least Median of Squares
NEW1	IERLS with unweighted final estimates
NEW2	IERLS with weighted final estimates
OLS	Ordinary Least Squares

The characteristics compared were: Efficiency relative to OLS; MSE ratio relative to OLS; bias defined as the absolute difference between true parameters and estimates from each procedure; and bias defined as the ratio of Efficiency to MSE ratio.

Efficiency

The variance of each estimate over 400 replications was calculated. Each estimated variance was then divided by its OLS counterpart, and the resultant value is the efficiency measure of the procedure producing the estimate. If this value is greater than 1, the procedure is less efficient than OLS. If it is less than 1, the procedure is more efficient than 1. If the value is 1, the procedure is as efficient as OLS. The following table gives the frequency counts of each of the six procedures when each is the best.

Huber's M-estimate with k=1.5 reduced the estimation variability the most in simple regression for both sample sizes. However, in multiple regression, HUB did not perform as well as it did in simple regression. In multiple regression with n=50, LMS was the most efficient procedure among all procedures used in the study. With n=100 and outlying dimensions dependent, LAD was the best. With n=100 and outlying dimensions independent, the IERLS with weighted average (NEW2) was the best. LMS was not used when sample size was 100.

Table 1 Frequency distributions of each of the six regression procedures with lowest variances of estimates

	simple regression		multiple regression (outlying dimension dependent)			multiple regression (outlying-dimension independent)		
	b0	b1	b0	b1	b2	b0	b1	b2
	f (%)	f (%)	f (%)	f (%)	f (%)	f (%)	f (%)	f (%)
n=50								
HUB	**25 (46)**	**22 (40)**	14 (11)	9 (7)	13 (10)	n a	n a	n a
LAD	9 (16)	8 (15)	15 (12)	18 (14)	17 (13)	n a	n a	n a
LMS	0 (0)	0 (0)	**54 (43)**	**66 (52)**	**59 (47)**	n a	n a	n a
NEW1	1 (2)	6 (15)	15 (12)	21 (17)	19 (15)	n a	n a	n a
NEW2	1 (2)	2 (4)	28 (22)	11 (9)	14 (11)	n a	n a	n a
OLS	19 (34)	17 (31)	1 (1)	2 (2)	5 (4)	n a	n a	n a
n=100								
HUB	**29 (53)**	**18 (33)**	5 (4)	0 (0)	10 (8)	16 (13)	13 (10)	25 (20)
LAD	7 (13)	11 (20)	**75 (59)**	**64 (50)**	**56 (44)**	41 (32)	35 (28)	**44 (35)**
NEW1	10 (18)	11 (20)	5 (4)	8 (6)	9 (7)	13 (10)	6 (5)	6 (8)
NEW2	4 (7)	0 (0)	36 (28)	42 (33)	38 (30)	**54 (43)**	**54 (43)**	42 (33)
OLS	5 (9)	15 (27)	6 (11)	21 (17)	19 (15)	3 (2)	19 (15)	10 (8)

It can be seen from the above table that under the most unfavorable situations where outlying dimensions are independent, IERLS reduced the estimation variability the most It can also be seen that IERLS with unweighted average, although somewhat simpler then its weighted average counterpart, performed less satisfactory with larger sample size and two predictors

Mean Squared Errors

The following table summarizes information of the performance in MSE reduction for all six procedures Huber's M-estimate with k=1 5 had the lowest mean squared errors among all procedures in simple regression, for both n=50 and n=100 OLS performed surprisingly well under the same situations However, with n=100 the performance of HUB and OLS deteriorate LMS excelled all in multiple regression with n=50 When n=100, IERLS with weighted average was the best for either dependent or independent

outlying dimensions LAD was also a relatively better procedure in reducing the values of MSE

Table 2 Frequency distributions of each of the six regression procedures with lowest MSE for each coefficient estimate

	simple regression		multiple regression			multiple regression		
	b0	b1	b0	b1	b2	b0	b1	b2
	f (%)	f (%)	f (%)	f (%)	f (%)	f (%)	f (%)	f (%)
n=50								
HUB	17 (31)	**21 (38)**	13 (10)	5 (4)	7 (6)	n a	n a	n a
LAD	15 (27)	9 (16)	16 (13)	14 (11)	15 (12)	n a	n a	n a
LMS	0 (0)	0 (0)	**55 (43)**	**71 (56)**	**71 (56)**	n a	n a	n a
NEW1	0 (0)	6 (11)	14 (11)	21 (17)	18 (14)	n a	n a	n a
NEW2	4 (7)	8 (15)	28 (22)	15 (12)	15 (12)	n a	n a	n a
OLS	**19 (35)**	17 (31)	1 (1)	1 (1)	1 (1)	n a	n a	n a
n=100								
HUB	**25 (46)**	**19 (35)**	10 (8)	0 (0)	4 (3)	10 (8)	6 (5)	12 (9)
LAD	7 (13)	10 (18)	**61 (48)**	49 (39)	48 (38)	41 (32)	29 (23)	33 (26)
NEW1	7 (13)	11 (20)	2 (2)	9 (7)	10 (8)	5 (4)	6 (5)	3 (2)
NEW2	6 (11)	1 (2)	53 (42)	**65 (51)**	**60 (47)**	**71 (56)**	**84 (66)**	**78 (61)**
OLS	10 (18)	14 (26)	1 (1)	4 (3)	5 (4)	0 (0)	2 (2)	1 (1)

However, NEW1 (IERLS with unweighted average) performs poorly compared to the other robust procedures

Bias

Bias in this study was measured from two perspectives First is the difference between the true parameters and the estimates produced by each of the procedures The smaller the difference, the smaller the bias Tables 3, 4, and 5 summarize frequency counts of each procedure when it was the best

Table 3 indicates that there was little consistency with respect to which procedure produced estimates closest to true parameters in simple regression When LMS was used it was the best in this aspect When it was not used, with n=50 the IERLS with weighted averages was the best, while with n=100 OLS was the best for b_0 and LAD for b_1 Table 4 shows that in multiple regression, with n=50 LMS was the best when used When LMS

Table 3 Frequency distributions of each of the six regression procedures with smallest differences between estimates and corresponding parameters in simple regression

	n=50				n=100	
	Simple regression (with LMS)		Simple regression (without LMS)		Simple regression (without LMS)	
	b0	b1	b0	b1	b0	b1
	f (%)	f (%)	f (%)	f (%)	f (%)	f (%)
HUB	11 (20)	3 (6)	11 (20)	3 (6)	18 (33)	0 (0)
LAD	10 (18)	5 (9)	19 (35)	5 (9)	8 (15)	**24 (44)**
LMS	**25 (46)**	**19 (35)**	n.a.	n.a.	n.a.	n.a.
NEW1	3 (6)	9 (16)	5 (9)	17 (31)	4 (7)	13 (24)
NEW2	6 (11)	14 (25)	**20 (36)**	**24 (43)**	5 (9)	17 (31)
OLS	0 (0)	5 (9)	0 (0)	6 (11)	**20 (36)**	1 (2)

Table 4 Frequency distributions of each of the six regression procedures with smallest differences between estimates and corresponding parameters in multiple regression (n=50)

	Multiple regression (with LMS)			Multiple regression (without LMS)		
	b0	b1	b2	b0	b1	b2
	f (%)	f (%)	f (%)	f (%)	f (%)	f (%)
HUB	6 (5)	0 (0)	0 (0)	12 (9)	0 (0)	0 (0)
LAD	12 (9)	5 (4)	3 (2)	22 (17)	15 (12)	16 (13)
LMS	**91 (72)**	**90 (71)**	**96 (76)**	n.a.	n.a.	n.a.
NEW1	10 (8)	8 (6)	8 (6)	21 (17)	15 (12)	15 (12)
NEW2	5 (4)	16 (13)	10 (8)	**66 (52)**	**88 (69)**	**82 (65)**
OLS	3 (2)	8 (6)	10 (8)	6 (5)	9 (7)	14 (11)

Table 5 Frequency distributions of each of the six regression procedures with smallest differences between estimates and corresponding parameters in multiple regression (n=100)

	Multiple regression (without LMS) (outlying dimension dependent)			Multiple regression (without LMS) (outlying dimension independent)		
	b0	b1	b2	b0	b1	b2
n=100	f (%)	f (%)	f (%)	f (%)	f (%)	f (%)
HUB	28 (22)	5 (4)	2 (2)	7 (6)	4 (3)	1 (1)
LAD	14 (11)	25 (20)	36 (28)	31 (24)	18 (14)	19 (15)
LMS	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
NEW1	11 (9)	21 (17)	20 (16)	20 (17)	16 (13)	17 (13)
NEW2	**67 (53)**	**67 (53)**	**66 (52)**	**64 (50)**	**85 (67)**	**86 (68)**
OLS	7 (5)	9 (7)	3 (2)	5 (4)	7 (6)	4 (3)

was not used, IERLS with weighted averages was the best. When n was 100, IERLS with weighted averages was the overwhelming winner in all three estimates (Table 5).

Another way of looking at bias is to compare the ratio of MSE ratio to efficiency of estimates from each procedure. The smaller the ratio, the more reduction in bias. Table 6 gives this information. LMS was the best among the five procedures whenever it was used. When LMS was not used, IERLS with weighted averages became the best one in reducing bias.

Table 6. Frequency distributions of each of the six regression procedures with lowest bias ratios of each coefficient estimate

	simple regression		multiple regression (outlying dimensions dependent)			multiple regression (outlying dimensions independent)		
	b0	b1	b0	b1	b2	b0	b1	b2
	f (%)	f (%)	f (%)	f (%)	f (%)	f (%)	f (%)	f (%)
n=50								
HUB	3 (6)	14 (26)	11 (9)	0 (0)	0 (0)	n.a	n.a	n.a
LAD	2 (4)	12 (22)	10 (8)	0 (0)	0 (0)	n.a	n.a	n.a
LMS	**36 (66)**	17 (31)	**87 (69)**	**65 (51)**	**73 (58)**	n.a	n.a	n.a
NEW1	1 (2)	5 (9)	11 (8)	3 (2)	9 (7)	n.a	n.a	n.a
NEW2	13 (24)	**32 (58)**	8 (6)	59 (47)	45 (35)	n.a	n.a	n.a
n=100								
HUB	15 (27)	1 (2)	21 (17)	9 (7)	7 (6)	3 (2)	12 (9)	6 (5)
LAD	**16 (29)**	21 (38)	8 (6)	8 (6)	9 (7)	26 (21)	14 (11)	28 (22)
NEW1	14 (26)	11 (20)	23 (18)	22 (17)	30 (24)	39 (31)	18 (14)	20 (16)
NEW2	10 (18)	**22 (40)**	**75 (59)**	**88 (69)**	**81 (64)**	**59 (47)**	**83 (65)**	**73 (58)**

Comparison of Execution Time of All Procedures

Data with large sample size could slow down the speed of the IERLS procedure, since it will need to perform $\frac{n}{2}$ iterations. Execution time (in seconds) for all six procedures with different sample sizes are listed in the following table (in actual seconds, not CPU time).

Table 7. Time comparison of the six regression procedures with different sample size in both simple and multiple regression with two predictors

	Simple regression						Multiple regression					
n	HUB	LAD	LMS	NEW1	NEW2	OLS	HUB	LAD	LMS	NEW1	NEW2	OLS
50	<0.01	<0.01	2.25	<0.01	<0.01	<0.01	<0.01	<0.01	3.57	<0.01	<0.01	<0.01
100	<0.01	<0.01	9.39	<0.01	0.05	<0.01	<0.01	<0.01	16.25	<0.01	0.06	<0.01
200	n.a.	n.a.	42.23	n.a.	n.a.	n.a.	n.a.	n.a.	187.24	n.a.	n.a.	n.a.
300	n.a.	n.a.	94.86	n.a.	n.a.	n.a.	n.a.	n.a.	-	n.a.	n.a.	n.a.
500	0.28	<0.01	-	0.65	0.87	<0.01	0.38	<0.01	-	0.99	1.26	<0.01
1000	1.27	<0.01	-	2.56	3.02	<0.01	0.99	<0.01	-	3.79	4.01	<0.01

Note that for LMS the sample size is limited to 300, the maximum capacity a 200 Mhz Pentium with 16 MB memory running the LMS program. However, if the program is executed in an MS-Windows platform (not MS-DOS), the Extended memory can probably solve this problem. Similarly, for all the other procedures, the largest sample size this type of computers can handle is 1000 (through trial-and-error by the researcher) Two sample sizes (200 and 300) were added to check the speed of LMS, since when sample size was 500 the LMS program would not run.

Based on this limited comparison on run speed among these programs, it is apparent that those that do not require iteration (i.e., LAD and OLS) take fewer time to execute, while the speed slows down with more and more iterations. The proposed iteratively reweighted least squares procedure with weighted averages (NEW2), however, does not seem to be lagging too much even with a sample size of 1000. The LMS program is significantly slower than the other four programs.

Summary

This simulation study attempted to compare the capacity of these robust procedures to recover the underlying model, with varying degrees of outlier contamination in the data, using two sample sizes (50 and 100), in both simple and multiple regression situations. The five robust techniques, as well as OLS, were compared in terms of the variance of estimates, the values of MSE and bias of the estimates. Although LMS appears to be the best, the high demand on computing resource would limit its applicability. OLS took the least time to solve, however, it can get highly unstable and inaccurate with outlier

contamination. The natural choice in this study seems to fall among LAD, HUB and NEW1 or NEW2.

From the simulation, it could be suggested that if the reduction of variance of estimates is very important, HUB should be used in simple regression. In multiple regression LMS is best in reducing variance. However, if there is reason to believe that outliers are "peculiar" in all directions, the proposed IERLS procedure (NEW2) should be used. If the reduction of bias is the main focus, then the IERLS procedure should also be used, since it was best in lowering the bias of estimation.

The magnitude of contamination used were: 3, 6, 9, and -3, -6, -9. It appears that the sign (+/-) did not have much impact on regression estimates when the correlation between variables is zero, however, the sign can have some impact when opposite to the correlation among variables. The absolute value of the magnitude represents the "strength" of the contamination. Such "strength" has an impact on the estimates. Generally speaking, as the value went up, the variance and bias went up as well. The variance ratio and MSE ratio both seem to be lowered as magnitudes became larger, using the robust procedures. In addition, those ratios were in general lowered when the percentage of contamination increased.

The results of comparisons among HUB, LAD, LMS, and NEW1 and NEW2 are mixed. No single one among these is consistently better the others. However, if the main goal is to obtain the estimates least influenced by outliers, IERLS seems to be more appropriate since NEW2 for the most part produced estimates closest to the true parameters (without considering LMS).

Although selection of the best procedure with respect to all three criteria: smallest variance, lowest MSE, and the least bias, seems difficult, it should be kept in mind that there are two major purposes in robust regression: first, to identify outliers, second, to estimate the regression parameters. If the correct identification of outliers is the major

concern, it seems that the procedure with the lowest MSE should be used, since it will give more consistent estimates and thus more consistent identification to "strange" data points. On the other hand, if the estimation of regression parameter is the major concern, then the procedure giving most accurate (or the least biased) estimates should be used. Since this study focuses on the estimation of regression parameters, the appropriate criterion would thus be the bias, or the MSE/Variance ratio. The proposed IERLS procedure thus seems to be a better choice.

Suggestions for Further Research

A major question concerns the standard error of the IERLS estimates. Although Lin and Davenport (1997) proposed a way to calculate them, they found the resulting values much liberal than the standard errors obtained from the simulation with 400 replications.

The calculation of standard errors of the IERLS estimates is complex. The coefficients obtained at each iteration were based on different number of data points, thus it is unlikely that each set of estimates would have homogeneous variance compared to the others. The closed form of IERLS standard errors remain to be discovered.

The statistical power of this procedure is also unknown. However, this could be easily done by calculating the "z-score" for each condition for each procedure across the 400 replications, and counting the ones that are significant at certain alpha level when the parameters are non-zero. The type I error rate, and type II error rate can be calculated in the same manner.

It is possible to apply this algorithm to the calculation of other parametric statistics from a mean, a correlation coefficient (even the analysis of variance), to regression coefficients as discussed. Further, this weighting algorithm could be applied to most GLM procedures, as well as discriminant analysis and factor analysis. In fact, it is believed that such algorithm should provide researchers a different view into the data set they are working on.

The magnitude of outlying was limited to no larger than |9| times the initial value. However, it would be interesting to see the results if larger magnitude is used. Also, the breakdown point of this procedure is not clear from this study and could be a point of interest.

A Concluding Remark

No matter how much advance has been made in robust regression, any robust procedure should not replace human judgment. In this study, all the models were so constructed that relationship between the dependent variable and the independent variable(s) was linear. If the model is incorrect, then the robust fit will actually result in regression estimates that are even further away from the truth compared with OLS. Therefore, it is always a good practice to examine the model first, before performing analysis with the data.

References

Banchoff, T., & Wermer, J. (1992). Linear Algebra Through Geometry. New York: Springer-Verlag.

Bloomfield, P., & Steiger, W. L. (1983). Least Absolute Deviations: Theory, Applications, and Algorithms. Boston: Birkhauser.

Chatterjee, S., & Hadi, A. (1988). Sensitivity Analysis in Linear Regression. New York: John Wiley & Sons, Inc.

Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. Annals of Statistics, 1, 799-821.

Lewis, P. A. W., & Orav, E. J. (1989). Simulation Methodology for Statisticians, Operations analysts, and Engineers, Volume 1. Belmont, CA: Wadsworth, Inc.

Lin, C., & Davenport, E. C., Jr. (1996). A weighted least squares approach to robustify least squares estimates. Paper presented at the annual meeting of the American Educational Research Association, March 1996, Chicago.

Rousseeuw, P. J. (1984). Least median of squares regression. Journal of the American statistical association, 79, 871-880.

Rousseeuw, P. J., & Leroy, A. M. (1987). Robust Regression and Outlier Detection. New York: John Wiley & Sons, Inc.

Tietjen, G. L. (1986). A Topical Dictionary of Statistics. New York: Chapman and Hall Ltd.

Woodward, J. A., & Overall, J. (1974). Computer generation of random normal deviates for Monte Carlo work. Perceptual and motor skills, 39, 31-37.

Appendix 1.
Illustration of the relationship between MSE and variances.

Define the following variables

VAR_A: the variance of estimates using procedure A
MSE_A: the Mean Squared Error using procedure A
BIS_A: the bias using procedure A
VAR_OLS: the variance of estimates using OLS
MSE_OLS: the Mean Squared Error of estimates using OLS
BIS_OLS: the bias using OLS

Further suppose that (where a, b are non-negative constants)

$$VAR_A = a \times VAR_OLS, \qquad (i)$$

and

$$MSE_A = b \times MSE_OLS \qquad (ii)$$

It is established that

$$MSE = VAR + BIS.$$

Thus,

$$MSE_OLS = VAR_OLS + BIS_OLS, \qquad (iii)$$

and

$$MSE_A = VAR_A + BIS_A \qquad (iv)$$

Applying (i) and (ii) into (iii) and (iv), one has

$$\begin{aligned} MSE_A &= VAR_A + BIS_A \\ &= a \times VAR_OLS + BIS_A \end{aligned}$$

Thus,

$$b \times MSE_OLS = a \times VAR_OLS + BIS_A$$

Furthermore,

$$BIS_OLS = MSE_OLS - VAR_OLS, \qquad (v)$$

and

$$BIS_A = b \times MSE_OLS - a \times VAR_OLS \qquad (vi)$$

It follows that $BIS_A \le BIAS_OLS$, provided that $b \le a$. The reduction in BIS_A will become more evident when the difference between a and b is larger.

Therefore, an increase in VAR_A could be a signal of a reduction in BIS_A. Divide b by a, if the resulting value is smaller than 1, then BIS_A is smaller than BIS_OLS, and using procedure A can yield lower bias.

In this current study, one can divide the MSE ratio by its corresponding efficiency, and compare the resulting value with 1 as described above.

Appendix 2.
Conditions Used for Simple Regression.

Condition	Direction	Magnitude	Percentage
111	X_1	×3	0
112	X_1	×3	5
113	X_1	×3	15
114	X_1	×3	25
121	X_1	×6	0
122	X_1	×6	5
123	X_1	×6	15
124	X_1	×6	25
131	X_1	×9	0
132	X_1	×9	5
133	X_1	×9	15
134	X_1	×9	25
141	X_1	×(-3)	0
142	X_1	×(-3)	5
143	X_1	×(-3)	15
144	X_1	×(-3)	25
151	X_1	×(-6)	0
152	X_1	×(-6)	5
153	X_1	×(-6)	15
154	X_1	×(-6)	25
161	X_1	×(-9)	0
162	X_1	×(-9)	5
163	X_1	×(-9)	15
164	X_1	×(-9)	25
211	X_2	×3	0
212	X_2	×3	5
213	X_2	×3	15
214	X_2	×3	25
221	X_2	×6	0
222	X_2	×6	5
223	X_2	×6	15
224	X_2	×6	25
231	X_2	×9	0
232	X_2	×9	5
233	X_2	×9	15
234	X_2	×9	25
241	X_2	×(-3)	0
242	X_2	×(-3)	5
243	X_2	×(-3)	15
244	X_2	×(-3)	25
251	X_2	×(-6)	0
252	X_2	×(-6)	5
253	X_2	×(-6)	15
254	X_2	×(-6)	25

Appendix 2. (Continued.)

Condition	Direction	Magnitude	Percentage
261	X_2	×(-9)	0
262	X_2	×(-9)	5
263	X_2	×(-9)	15
264	X_2	×(-9)	25
311	X_1, X_2	×3	0
312	X_1, X_2	×3	5
313	X_1, X_2	×3	15
314	X_1, X_2	×3	25
321	X_1, X_2	×6	0
322	X_1, X_2	×6	5
323	X_1, X_2	×6	15
324	X_1, X_2	×6	25
331	X_1, X_2	×9	0
332	X_1, X_2	×9	5
333	X_1, X_2	×9	15
334	X_1, X_2	×9	25
341	X_1, X_2	×(-3)	0
342	X_1, X_2	×(-3)	5
343	X_1, X_2	×(-3)	15
344	X_1, X_2	×(-3)	25
351	X_1, X_2	×(-6)	0
352	X_1, X_2	×(-6)	5
353	X_1, X_2	×(-6)	15
354	X_1, X_2	×(-6)	25
361	X_1, X_2	×(-9)	0
362	X_1, X_2	×(-9)	5
363	X_1, X_2	×(-9)	15
364	X_1, X_2	×(-9)	25

Appendix 3.
Conditions Used for Multiple Regression

Condition	Direction	Magnitude	Percentage
111	X_1	×3	0
112	X_1	×3	5
113	X_1	×3	15
114	X_1	×3	25
121	X_1	×6	0
122	X_1	×6	5
123	X_1	×6	15
124	X_1	×6	25
131	X_1	×9	0
132	X_1	×9	5
133	X_1	×9	15
134	X_1	×9	25
141	X_1	×(-3)	0
142	X_1	×(-3)	5
143	X_1	×(-3)	15
144	X_1	×(-3)	25
151	X_1	×(-6)	0
152	X_1	×(-6)	5
153	X_1	×(-6)	15
154	X_1	×(-6)	25
161	X_1	×(-9)	0
162	X_1	×(-9)	5
163	X_1	×(-9)	15
164	X_1	×(-9)	25
211	X_2	×3	0
212	X_2	×3	5
213	X_2	×3	15
214	X_2	×3	25
221	X_2	×6	0
222	X_2	×6	5
223	X_2	×6	15
224	X_2	×6	25
231	X_2	×9	0
232	X_2	×9	5
233	X_2	×9	15
234	X_2	×9	25
241	X_2	×(-3)	0
242	X_2	×(-3)	5
243	X_2	×(-3)	15
244	X_2	×(-3)	25
251	X_2	×(-6)	0
252	X_2	×(-6)	5
253	X_2	×(-6)	15
254	X_2	×(-6)	25

Appendix 3. (Continued)

Condition	Direction	Magnitude	Percentage
261	X_2	×(-9)	0
262	X_2	×(-9)	5
263	X_2	×(-9)	15
264	X_2	×(-9)	25
311	Y	×3	0
312	Y	×3	5
313	Y	×3	15
314	Y	×3	25
321	Y	×6	0
322	Y	×6	5
323	Y	×6	15
324	Y	×6	25
331	Y	×9	0
332	Y	×9	5
333	Y	×9	15
334	Y	×9	25
341	Y	×(-3)	0
342	Y	×(-3)	5
343	Y	×(-3)	15
344	Y	×(-3)	25
351	Y	×(-6)	0
352	Y	×(-6)	5
353	Y	×(-6)	15
354	Y	×(-6)	25
361	Y	×(-9)	0
362	Y	×(-9)	5
363	Y	×(-9)	15
364	Y	×(-9)	25
411	X_1, X_2	×3	0
412	X_1, X_2	×3	5
413	X_1, X_2	×3	15
414	X_1, X_2	×3	25
421	X_1, X_2	×6	0
422	X_1, X_2	×6	5
423	X_1, X_2	×6	15
424	X_1, X_2	×6	25
431	X_1, X_2	×9	0
432	X_1, X_2	×9	5
433	X_1, X_2	×9	15
434	X_1, X_2	×9	25

Appendix 3. (Continued.)

Condition	Direction	Magnitude	Percentage
441	X_1, X_2	×(-3)	0
442	X_1, X_2	×(-3)	5
443	X_1, X_2	×(-3)	15
444	X_1, X_2	×(-3)	25
451	X_1, X_2	×(-6)	0
452	X_1, X_2	×(-6)	5
453	X_1, X_2	×(-6)	15
454	X_1, X_2	×(-6)	25
461	X_1, X_2	×(-9)	0
462	X_1, X_2	×(-9)	5
463	X_1, X_2	×(-9)	15
464	X_1, X_2	×(-9)	25
511	X_1, Y	×3	0
512	X_1, Y	×3	5
513	X_1, Y	×3	15
514	X_1, Y	×3	25
521	X_1, Y	×6	0
522	X_1, Y	×6	5
523	X_1, Y	×6	15
524	X_1, Y	×6	25
531	X_1, Y	×9	0
532	X_1, Y	×9	5
533	X_1, Y	×9	15
534	X_1, Y	×9	25
541	X_1, Y	×(-3)	0
542	X_1, Y	×(-3)	5
543	X_1, Y	×(-3)	15
544	X_1, Y	×(-3)	25
551	X_1, Y	×(-6)	0
552	X_1, Y	×(-6)	5
553	X_1, Y	×(-6)	15
554	X_1, Y	×(-6)	25
561	X_1, Y	×(-9)	0
562	X_1, Y	×(-9)	5
563	X_1, Y	×(-9)	15
564	X_1, Y	×(-9)	25
611	X_2, Y	×3	0
612	X_2, Y	×3	5
613	X_2, Y	×3	15
614	X_2, Y	×3	25
621	X_2, Y	×6	0
622	X_2, Y	×6	5
623	X_2, Y	×6	15
624	X_2, Y	×6	25

Appendix 3. (Continued.)

Condition	Direction	Magnitude	Percentage
631	X_2,Y	×9	0
632	X_2,Y	×9	5
633	X_2,Y	×9	15
634	X_2,Y	×9	25
641	X_2,Y	×(-3)	0
642	X_2,Y	×(-3)	5
643	X_2,Y	×(-3)	15
644	X_2,Y	×(-3)	25
651	X_2,Y	×(-6)	0
652	X_2,Y	×(-6)	5
653	X_2,Y	×(-6)	15
654	X_2,Y	×(-6)	25
661	X_2,Y	×(-9)	0
662	X_2,Y	×(-9)	5
663	X_2,Y	×(-9)	15
664	X_2,Y	×(-9)	25
711	X_1, X_2, Y	×3	0
712	X_1, X_2, Y	×3	5
713	X_1, X_2, Y	×3	15
714	X_1, X_2, Y	×3	25
721	X_1, X_2, Y	×6	0
722	X_1, X_2, Y	×6	5
723	X_1, X_2, Y	×6	15
724	X_1, X_2, Y	×6	25
731	X_1, X_2, Y	×9	0
732	X_1, X_2, Y	×9	5
733	X_1, X_2, Y	×9	15
734	X_1, X_2, Y	×9	25
741	X_1, X_2, Y	×(-3)	0
742	X_1, X_2, Y	×(-3)	5
743	X_1, X_2, Y	×(-3)	15
744	X_1, X_2, Y	×(-3)	25
751	X_1, X_2, Y	×(-6)	0
752	X_1, X_2, Y	×(-6)	5
753	X_1, X_2, Y	×(-6)	15
754	X_1, X_2, Y	×(-6)	25
761	X_1, X_2, Y	×(-9)	0
762	X_1, X_2, Y	×(-9)	5
763	X_1, X_2, Y	×(-9)	15
764	X_1, X_2, Y	×(-9)	25

Appendix 4.
Parameter Estimates Studied in Simulation.

Parameter Estimates	Simple Regression	Multiple Regression
$\hat{\beta}_0$	Yes	Yes
$\hat{\delta}^2_{\beta_0}$	Yes	Yes
$BIAS_{\beta_0}$	Yes	Yes
$\hat{\beta}_1$	Yes	Yes
$\hat{\delta}^2_{\beta_1}$	Yes	Yes
$BIAS_{\beta_1}$	Yes	Yes
$\hat{\beta}_2$	N A	Yes
$\hat{\delta}^2_{\beta_2}$	N A	Yes
$BIAS_{\beta_2}$	N A	Yes

Appendix 5.
Annotated SPSS Program for IERLS.

```
* Lines beginning with an "*" are comment lines.
* Lines in boldface are command lines.
* For Windows 3.1 version, the period after each command is necessary.

set mxloop=100.
* For larger sample size, a larger mxloop number may be needed.

matrix.

*A is the active matrix storing data, AXY.SAV the SPSS data file,
*with intercept (1's) in the first column and dependent variable
*in the last column.

get A/file='c:AXY.sav'.

compute nr=nrow(A).
compute nc=ncol(A).

compute ATP=make(nr,nc-1,0).

*skip the firt column of 1's.

loop i=2 to nc.
    compute avg=csum(A(:,i))/nr.
    print {i,avg}.
    loop j=1 to nr.
        compute ATP(j,i-1)=A(j,i)-avg.
    end loop.
end loop.

compute M=SSCP(ATP)/(nr-1).

*Uncommmenting the next line helps verify the correctness of the
*variance/covariance matrix, M.

*print M.

*DT is the matrix holding distance value for each data point.

compute DT=make(nr,1,1).
loop j=1 to nr.
    compute DT(j,1)=ATP(j,:)*inv(M)*t(ATP(j,:)).
end loop.

*Sort data points on magnitude of DT.

loop j=1 to nr-1.
   loop k=j+1 to nr.
      do if(DT(j,1)>DT(k,1)).
        compute TMP=DT(j,1).
        compute DT(j,1)=DT(k,1).
        compute DT(k,1)=TMP.
```

```
        loop i=1 to nc.
            compute TMP=A(j,i).
            compute A(j,i)=A(k,i).
            compute A(k,i)=TMP.
        end loop.
      end if.
    end loop.
end loop.
*creating space to hold regression coefficients, and to
*accumulate necessary values for calculating standard errors
*of the coefficients.

compute nnc=nc-1.
compute BETA=make(nnc,1,-100).
compute NBETA=make(nnc,1,0).
compute N2BETA=make(nnc,1,0).

*Determining the number of data points for the initial basis.

do if(mod(nr,2)=0).
   compute nr2=nr/2.
else if(mod(nr,2)>0).
   compute nr2=(nr+1)/2.
end if.
print nr2.
compute ATP=make(nr2,nnc,1).
computw ATP=A(1:nr2,1:nnc).
print ATP.

*calculating coefficients.

compute BETA=inv(t(ATP)*ATP)*t(ATP)*A(1:nr2,nc).

*Creating space to hold weights.

compute WGT=make(nr,1,-1000).

*Calculating absolute values of residuals.

loop i=1 to nr.
  compute TMP=0.
  loop j=2 to nnc.
   compute TMP=BETA(j,1)*A(i,j)+TMP.
  end loop.
  compute TMP=TMP+BETA(1,1).
  compute TMP=A(i,nc)-TMP.
  compute WGT(i,1)=abs(TMP).
end loop.

*Sort data points on magnitude of residuals.

loop i=1 to nr-1.
  loop i2=i+1 to nr.
    do if(WGT(i,1)>WGT(i2,1)).
        loop j=1 to nc.
          compute TMP=A(i,j).
```

```
            compute A(i,j)=A(i2,j).
            compute A(i2,j)=TMP.
          end loop.
          compute TMP=WGT(i,1).
          compute WGT(i,1)=WGT(i2,1).
          compute WGT(i2,1)=TMP.
        end if.
      end loop.
end loop.

*Accumulating weighted sum and weighted uncorrect SS for the
*regression coefficients.

compute AVG=make(nnc,1,0).
compute AVG=AVG+nr2*BETA.
compute TT1=rssq(BETA).
compute N2BETA=TT1*nr2 + N2BETA.

*Initializing weight to accumulate later, to calculate overall
*regression coefficients.

compute ICOUNT=nr2.

*Assigning weight to individual data points to calculate regression
*coefficients in each iteration.  Should not be confused with ICOUNT.

loop i=1 to nr.
   do if(i<=nr2).
      compute WGT(i,1)=1.
   else if(i>nr2).
      compute WGT(i,1)=1/(1+WGT(i,1)).
   end if.
end loop.

loop kk=nr2+1 to nr.

*Rescaling weight for each data point such that the sum of weights is
*the number of data points used in each iteration.

  compute ICOUNT=ICOUNT+kk.
  compute TMP=0.
  loop i=1 to kk.
    compute TMP=TMP+WGT(i,1).
  end loop.
  loop i=1 to kk.
   compute WGT(i,1)=sqrt(WGT(i,1)*kk/TMP).
   loop j=1 to nc.
     compute A(i,j)=A(i,j)*WGT(i,1).
   end loop.
  end loop.

*Calculating the regression coefficient at each kk.

compute ATP=make(kk,nnc,1).
compute ATP=A(1:kk,1:nnc).
compute BETA=inv(t(ATP)*ATP)*t(ATP)*A(1:kk,nc).
```

```
*Accumulating sum, and UNcorrected SS of regression coefficients.

compute AVG=AVG+kk*BETA.
compute TT1=rssq(BETA).
compute N2BETA=N2BETA+kk*TT1.
end loop.

*Calculating square of sum, and standard errors for the regression
*coefficients.

compute NBETA=rssq(AVG).
compute NBETA=NBETA/ICOUNT.
compute AVG=AVG/ICOUNT.
compute NBETA=sqrt((N2BETA-NBETA)/(ICOUNT-1)).

*AVG is the column vector of overall weighted regression coefficients
*NBETA is the column vector holding standard errors of coefficients
*in AVG.

print AVG.
print NBETA.
end matrix.
```